

NPS55-85-021

NAVAL POSTGRADUATE SCHOOL

Monterey, California



TECHNICAL

ESTIMATION OF SEA-SURFACE WINDSPEED FROM
WHITECAP COVER: STATISTICAL APPROACHES
COMPARED EMPIRICALLY AND BY SIMULATION

by

I. G. O'Muircheartaigh
D. P. Gaver

September 1985

Approved for public release; distribution unlimited.

Prepared for: Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

FEDDOCS
D 208.14/2
NPS-55-85-021

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral R. H. Shumaker
Superintendent

David A. Schrady
Provost

Reproduction of all or part of this report is authorized.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55-85-021	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER NAVAL POSTGRADUATE SCHOOL MONTEREY CA 93943-5101
4. TITLE (and Subtitle) ESTIMATION OF SEA-SURFACE WINDSPEED FROM WHITECAP COVER: STATISTICAL APPROACHES COMPARED EMPIRICALLY AND BY SIMULATION		5. TYPE OF REPORT & PERIOD COVERED Technical Report
7. AUTHOR(s) I. G. O'Muircheartaigh D. P. Gaver		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93943-5100		8. CONTRACT OR GRANT NUMBER(s)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 North Quincy Street Arlington, VA 22217		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N; RR014-05-01 N0001485WR24061
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE September 1985
		13. NUMBER OF PAGES 55
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) meteorology, whitecaps, statistics, remote sensing		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Several (5) statistical estimators for solving the typical remote sensing "inverse" problem are explained, illustrated and evaluated using both real data sets connecting surface windspeed and whitecap coverage and simulated data. The results suggest the superiority of some of these estimators.		

ESTIMATION OF SEA-SURFACE WINDSPEED FROM WHITECAP COVER:

STATISTICAL APPROACHES COMPARED EMPIRICALLY AND BY SIMULATION

by

I.G. O'MUIRCHEARTAIGH
University College
Galway, Ireland

D.P. GAVER
Naval Postgraduate School
Monterey, CA 93943

1. INTRODUCTION

It is well known that the extent of whitecap cover on the surface of a sea is greatly influenced by the surface windspeed (Monahan (1971), Toba and Chaen (1973), Wu (1979), Monahan and O'Muircheartaigh (1980)). Other variables, such as sea surface temperature, also are important, but windspeed action appears to play the dominant role. Whitecap cover can be remotely sensed while windspeed cannot, so it is tempting to utilize the relationship between windspeed and whitecaps to infer reasonable values for the surface windspeed. To do so requires that the natural causative relation of "whitecaps windspeed", quantitatively estimated from field data as a statistical regression of (some measure of) white cap coverage on windspeed, be reversed. It turns out that "natural" way of solving the problem, namely by regressing whitecap cover on windspeed and then inverting that regression relation, actually produces results that are inferior to those from some other procedures. Since the indirect remote sensing of windspeed is of operational interest, and since similar problems may well arise in different remote sensing, and other, areas we present

illustrative statistical data analyses of several sets of whitecaps-wind speed data in this paper. We also include, in later sections of the paper, further similar analyses based on simulated data.

The general problem considered here is that of making inferences about an unknown $p \times 1$ vector \underline{X}' from a single random observed $q \times 1$ response vector \underline{Y}' . The relationship between \underline{Y} and \underline{X} is calibrated with experimental data $(\underline{Y}_i, \underline{X}_i)$, $i = 1, 2, \dots, n$ where \underline{Y}_i , \underline{X}_i are $q \times 1$ and $p \times 1$ vectors, respectively. The case $p = q = 1$ has been extensively discussed in the literature, and reference will be made below to several basic contributions to calibration methods for this case. The situation when at least one of p, q is greater than one is the subject of a comprehensive paper by Brown (1982).

Brown (1982) distinguishes two cases of interest: (a) when both \underline{X} and \underline{Y} are random and (b) when only \underline{Y} is random, and \underline{X} can be fixed at prechosen levels. The former case is called random calibration, and the latter controlled calibration. The present paper is concerned solely with the problem of random calibration, because the data of interest arises in an observational context, and not from a controlled experiment.

A brief outline of the paper is as follows: In Section 2 we describe several different plausible methods of point estimation in univariate calibration. The methods described are subsequently applied to four data sets, and their performance evaluated in Section 3. In Section 4 we consider four interval estimates associated with the calibration problem, and apply them to the data sets. The problem of multivariate calibration is examined in Section 5. Several of the univariate methods are extended to this situation and applied to the same four data sets, and to a further set

provided in Brown (1982). The application and an evaluation of the results are presented in Section 6.

The later sections of the paper consider the same problems, but in the context of a simulation study. Section 7 gives a brief description of the objectives of the simulation study, Section 8 describes the point estimation results, and Section 9 those related to interval estimation.

2. THE UNIVARIATE PROBLEM

The simplest version of the calibration problem, and the one most extensively discussed, is the case $p = q = 1$, and where the calibration curve is linear in both the parameters and the independent variable. The situation of interest may therefore be described as follows: given two random variables X, Y with the relationship

$$Y = \alpha + \beta X + \epsilon \quad (2.1)$$

where, most classically $\epsilon \sim N(0, \sigma^2)$, and given n independent pairs of observations (X_i, Y_i) on (X, Y) and a new observation y_0 on Y , how do we predict or estimate the corresponding value of $X = X(y_0)$. Numerous solutions have been proposed, and their performances evaluated. Five of these methods, in particular, have been applied in Section 3 to four data sets, that relate whitecap cover to surface windspeed. The four methods examined are these:

(i) the so-called classical method viz., estimate α, β in equation (2.1) by least squares, and then for $Y = y_0$ the predicted value of X is

$$\hat{X}_C = \frac{y_0 - \hat{\alpha}}{\hat{\beta}}, \quad \hat{\beta} \neq 0. \quad (2.2)$$

(ii) Krutchkoff (1967) suggested another estimator obtained by rewriting (2.1) as

$$X = \gamma + \delta Y + e \quad (2.3)$$

and obtaining least-squares estimators $\hat{\gamma}$, $\hat{\delta}$ of γ , δ ; the predicted value of X , given $Y = y_0$ will then be

$$\hat{X}_I = \hat{\gamma} + \hat{\delta}y_0,$$

so denoted because it is known as the inverse estimator.

Krutchkoff (1967) concluded by means of a Monte Carlo study that \hat{X}_I had uniformly smaller mean squared-error (MSE) than the classical estimator \hat{X}_C . In a later (1969) paper he concluded that this result was valid only within the calibration range, whereas, in fact, the reverse result held outside that range. Williams (1969) pointed out that for finite samples the MSE of the classical estimator was infinite and that of the inverse estimator finite, thus the use of the MSE for comparing these estimators is unsatisfactory.

(iii). Lwin & Maritz (1980) proposed an alternative estimator based on the fact that for this particular problem, the predictor of X_0 given by

$$\hat{X}^*(y_0) = E\{X|y=y_0\} \quad (2.4)$$

has minimum mean-squared error (provided σ and α , β are all known). By using consistent estimators of σ , α , and β and by approximating the marginal distribution of X with the corresponding empirical distribution function, Lwin & Maritz showed that the estimator

$$\hat{X}_E = \frac{\sum_{i=1}^n x_i f\{(y_0 - \hat{\alpha} - \hat{\beta}x_i)/\hat{\sigma}\}}{\sum_{i=1}^n f\left(\frac{y_0 - \hat{\alpha} - \hat{\beta}x_i}{\hat{\sigma}}\right)} \quad (2.5)$$

will, subject to easily satisfied regularity conditions, tend to the optimal estimator $\hat{X}^*(y_0)$ in mean square, where f is the error density function (presumed known; otherwise estimated).

(iv) A Bayesian methodology was introduced by Aitchinson & Dunsmore (1975). This method involves the assumptions that

(a) X, Y are Normal,

(b) $Y \sim N(\alpha + \beta x, \sigma^2)$

From these assumptions, it can be shown that the predictive distribution for X_0 , given n pairs of observations (X_i, Y_i) and a single new observation y_0 is proportional to

$$St\left\{n-1, \bar{x}, \left(1+\frac{1}{n}\right) \frac{\sum (x_i - \bar{x})^2}{n-1}\right\} St\left\{n-2, m, \left(1+\frac{1}{k}\right) \frac{v}{v}\right\} \quad (2.6)$$

where

$$\frac{1}{k} = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}, \quad m = \bar{y} - \hat{\beta}(x_0 - \bar{x})$$

$$v = n-2$$

and

$$v = S_{yy} - \hat{\beta}^2 S_{xx}$$

and $St\{k,b,c\}$ is the usual non-central Student's t-distribution with density function given by

$$f(u;c,b,k) = \frac{1}{Be(\frac{1}{2}, \frac{1}{2}k)(kc)^{1/2} \{1+(kc)^{-1}(u-b)^2\}^{(k+1)/2}} \quad (2.7)$$

The constant of proportionality in (2.6) must be obtained by numerical integration. The predictive distribution of (2.6) enables us to obtain either point or interval estimates of X_0 . The point estimates examined are (a) mean of predictive distribution, \hat{X}_{ME} and (b) mode of predictive distribution, \hat{X}_{MO} .

We have, therefore, five different estimators to be compared:

- (i) the classical predictor \hat{X}_C

- (ii) the inverse predictor \hat{X}_I
- (iii) the empirical predictor \hat{X}_E
- (iv) the mean of the predictive distribution \hat{X}_{ME}
- (v) the mode of the predictive distribution \hat{X}_{MO}

3. COMPARISON OF UNIVARIATE PREDICTORS

3.1 The Data

The five predictors were compared by applying them to four data sets. The data sets consist of measurements of instantaneous oceanic whitecap coverage (Y) and wind speed (X), and the object of the exercise is the prediction of X_0 given a new observation Y_0 . An initial inspection of the data suggested lognormal distributions for both X and Y and a log transformation gave an acceptable fit to a Normal distribution. Data points for which whitecap coverage was 0.0 were excluded from the analysis for several reasons, but particularly because it seemed reasonable to assume that a zero whitecap coverage gave no additional information in relation to wind speed over and above the conditional distribution of wind speed given zero whitecap coverage. The data sets involved were the following:

Data set 1: Monahan (1971)

Data set 2: Toba & Chaen (1973)

Data set 3: JASIN experiment (1978), (Monahan et al. (1981))

Data set 4: Strex experiment (1981), (Monahan et al. (1981))

The number of (pairs of) non-zero observations in the respective data sets were 43, 18, 37 and 78.

3.2 Method of Comparison of Estimators

For each data set, we excluded one data point at a time and obtained each of the five estimators based on the remaining data. We then predicted the x-value of the excluded point, given the y-value of that point, using each of the five estimators. This provided five predicted x-values for each point in each data set. Finally, for each of the five estimators and for each data set, we calculated the mean bias (MB) and the mean-squared prediction error (MSPE) defined as follows for a given data set:

$$MB = \sum (x_i - \hat{x}_i) / n \quad (3.1)$$

$$MSPE = \sum (x_i - \hat{x}_i)^2 / n \quad (3.2)$$

where n is the number of points in the data set.

3.3 Results

The results are presented in Tables 1 and 2.

TABLE 1

Bias of Estimators

Data Set	Estimator				
	\hat{x}_C	\hat{x}_I	\hat{x}_E	\hat{x}_{ME}	\hat{x}_{MO}
1	.0150	.0038	-.0039	.1969	-.0050
2	.0119	-.0068	-.0141	-.1110	-.0092
3	.0055	.0019	-.0151	.2831	.0080
4	-.0014	.0004	.0030	.0415	.0006

Table 1 shows that, in terms of bias, the estimator \hat{x}_{ME} (i.e., the mean of the predictive distribution of x) is uniformly the worst of the five estimators and the inverse estimator \hat{x}_I almost uniformly the least biased. The estimator \hat{x}_{MO} is close to but slightly worse than, \hat{x}_I in terms of bias. A two-way analysis of variance applied to the data of Table 1 yielded the obvious results in terms of significance.

TABLE 2
Mean-Squared Prediction Error

Data Set	Estimator				
	\hat{x}_C	\hat{x}_I	\hat{x}_E	\hat{x}_{ME}	\hat{x}_{MO}
1	.192	.059	.056	.060	.060
2	.205	.082	.086	.082	.083
3	.103	.066	.060	.067	.067
4	.149	.061	.062	.061	.061

This table shows that in terms of average squared prediction error, the classical predictor is once again uniformly the poorest, having mean prediction error in the range 2 to 3 times that of any of the other estimators. The remaining four estimators are very close in terms of predictive capacity for those data sets, with none uniformly better than the others. Once again, a two-way ANOVA yielded the expected results.

One advantage of the Aitchison and Dunsmore method is that it produces, in addition to the point predictions, the predictive distribution of x given $y = y_0$. From this it is possible to obtain shortest $100(\alpha\%)$ confidence intervals for x given $y = y_0$.

4. INTERVAL ESTIMATION FOR THE WIND/WHITECAP DATA

For each predicted value, and for each data set, the following 95% confidence intervals have been constructed.

- I1. The standard 95% confidence interval based on the inverse regression--i.e., regression of U on W.
- I2. An interval based on the Lwin & Maritz estimator, and using the standard deviation of the closely related estimator E8, derived in Section 8.
- I3. An interval based on the classical estimator \hat{x}_C , and described in Brownlee (1965)

The results are presented below:

TABLE 3

Confidence Intervals

Data Set	I1		I2		I3	
	% cov	Av length	% cov	Av length	% cov	Av length
1	97.7	.90	97.7	.89	96.3	1.72
2	96.6	1.04	94.6	1.01	95.6	2.23
3	89.2	.96	94.6	.95	90.3	1.84
4	93.5	.96	92.3	.96	94.2	1.62

In general, intervals I1 and I2 are very comparable. The analysis was performed, as in the case of point estimation, by omitting each point in turn and constructing a confidence interval based on an analysis of all the remaining points of the data set.

5. THE MULTIVARIATE CASE

Brown (1982) has studied the case $p > q$, not both 1, in some depth. While most of his analysis relates to the case of controlled calibration (i.e., \underline{X} not random), he does devote some attention to the random calibration situation. The model employed is a generalization of (2.1), viz.,

$$\underline{Y} = 1\underline{\alpha}^T + \underline{X} \underline{B} + \underline{E} \quad (5.1)$$

where \underline{Y} ($n \times q$), \underline{E} ($n \times q$) and \underline{X} ($n \times p$) are random matrices, and \underline{E} is a disturbance matrix from $N_q(\underline{0}, \underline{\Sigma})$. If units of \underline{X} and \underline{Y} are chosen so that the variables are post hoc centered at zero, we can, without loss of generality, rewrite equation 5.1 so that the constant term disappears and hence we have

$$\underline{Y} = \underline{X} \underline{B} + \underline{E} \quad (5.2)$$

Brown (1982) suggests three estimators for the multivariate situation. These are analogous to the predictors \hat{X}_C , \hat{X}_I and \hat{X}_E of Section 2 and are derived as follows:

I. From regression of \underline{Y} on \underline{X} (denoted L by Brown), and analagous to \hat{X}_C .

$$\hat{\underline{X}}_L = (\hat{\underline{B}} \hat{\underline{S}}^{-1} \hat{\underline{B}})^{-1} \hat{\underline{B}} \hat{\underline{S}}^{-1} \underline{y}_0 \quad (5.3)$$

where \underline{y}_0 is the newly observed single value of \underline{Y} which we are to use to predict \underline{X} , and $\hat{\underline{B}}$, $\hat{\underline{S}}$ are the usual least squares estimators of \underline{B} , $\underline{\Sigma}$ (Mardia, et al., 1979). Note that if we replace $\hat{\underline{B}}$, $\hat{\underline{S}}$ by their univariate counterparts, and putting $\hat{\alpha} = 0$ (following centering of the data) equation (5.3) does, as expected, reduce to equation (2.2).

The analysis which produces $\hat{\underline{X}}_C$ here performs a multivariate regression of \underline{Y} on \underline{X} . Brown (1982) suggests an alternative predictor \underline{X}_L , where in attempting to predict a component of \underline{X} (say X_j) we regress \underline{Y} on X_j alone, and obtain $\hat{\underline{X}}_L$, by a formula analogous to (5.3).

II. From multivariate regression of \underline{X} on \underline{Y} (denoted LB in Brown (1982))

$$\hat{\underline{X}}_{LB} = \underline{y}_0'(\underline{Y}'\underline{Y})^{-1}\underline{Y}'\underline{X} \quad (5.4)$$

Note that in this case, each component of \underline{X} is predicted ignoring all the other components of \underline{X} --in effect we carry out a multiple (not a multivariate) regression of each component of \underline{X} on \underline{Y} .

III. A generalization of the empirical method of Lwin & Maritz (denoted E in Brown (1982)). The extension is straightforward. Like (L) it uses the parametric regression of \underline{Y} on \underline{X} and derives that of \underline{X} on \underline{Y} by means of the empirical distribution of \underline{X} . Specifically, if \underline{y}_0' is a new $(q \times 1)$ observation on \underline{Y} the prediction for the corresponding \underline{X}' $(p \times 1)$ is

$$\hat{\underline{X}}_E = \sum w_i \underline{X}_i' \quad (5.5)$$

where \underline{X}_i' is the i th observation on \underline{X} and

$$w_i = f(\underline{Y}' | \underline{X}_i) / \sum_{i=1}^n f(\underline{Y}' | \underline{X}_i) \quad (5.6)$$

In the case of our analysis, f was assumed to be the multivariate normal regression density (Mardia et al. (1979), ch. 6) with parameters fixed at their least squares values. It is, of course, also possible to obtain the estimator \hat{X}_E for the problem of predicting each component of \underline{X} separately, given \underline{y}_0 . This estimator is denoted by $\hat{X}_{E'}$, following Brown (1982).

The above 5 predictors were applied to five data sets, constituted as follows:

(a) the four data sets of Section 2, each augmented by the inclusion of additional x -variables, viz., surface water temperature, and air temperature [i.e., $q = 1$, $p = 3$].

(b) the data set provided in Brown (1982), Section 4, relating four infrared reflectance responses of wheat (\underline{Y}) to determination of percent water, X_1 , and percent protein, X_2 , of the wheat [i.e., $q = 4$, $p = 2$].

The various predictors were compared using the same criteria as in Section 2, viz., the mean-squared prediction error where one point at a time is omitted, and then the \underline{x} -value for that point is predicted using all the remaining points for estimation purposes.

6. ANALYSIS OF RESULTS FOR MULTIVARIATE CASE

We consider the results for the Brown data and the Wind/Whitecap data separately. In Table 3 we present the results for the data of Section 3.

TABLE 3
Mean Squared Prediction Error

Data Set	Predictor					
	\hat{X}_L	$\hat{X}_{L'}$	\hat{X}_{LB}	\hat{X}_E	$\hat{X}_{E'}$	\hat{X}_{LB*}
1	.095	.192	.059	.059	.056	.040
2	.550	.205	.082	.079	.086	.110
3	.114	.103	.066	.061	.060	.072
4	.148	.149	.061	.062	.062	.056

Before comparing these predictors, a number of points should be noted.

(i) \hat{X}_L is simply the classical estimator when only wind and whitecap variables are taken into account so that this is identical with the estimator \hat{X}_C of Section 2.

(ii) By definition, \hat{X}_{LB} predicts each component of X separately and hence this also is the univariate \hat{X}_I (since \underline{Y} has only one component here).

(iii) Included in column 6 of Table 3 is the predictor \hat{X}_{LB*} , obtained simply by regressing the wind variable on all other variables in the analysis [whether \underline{X} or \underline{Y}].

A comparison of the columns of Table 3 reveals that none of the multivariate methods used leads to any noticeable improvement in terms of

predictive capacity over the "best" univariate predictor $\hat{X}_I(\hat{X}_{LB})$. Among the truly multivariate of these methods $[\hat{X}_L, \hat{X}_E]$, the empirical \hat{X}_E holds up extremely well, whereas the classical multivariate again is uniformly the worst.

In Table 4 we present the results for the Brown data:

TABLE 4

Mean Squared Prediction Error

Method	→	L	L'	E	E'	LB
Variable						
↓						
X_1		.003	.003	.031	.017	.003
X_2		.041	.041	.298	.076	.041

A comparison of the columns of Table 4 confirms the result of Brown (1982) that the methods L, LB are virtually indistinguishable in terms of predictive performance for this data set. This is at variance with all previous univariate results, and with the multivariate conclusions for the wind/whitecap data. As printed out in Brown (1982), these results should be treated with some caution, as the data are perhaps atypical in that such a large percent of the variation is explained by the model. (Brown predicted the x-values of 5 points using the remaining 16, and found for methods L, LB in all cases over 98% of variation explained by the model. A similar

analysis by us for 15 other random samples of size 5 yielded an average of just under 98% of variation explained.)

Another interesting outcome of this analysis is the relatively poor performance of the method E for this data set. Our results confirm those of Brown, and in fact indicate that E is worse than in his analysis. Incidentally, an examination of the w_i (weights) involved in method E reveals that when we go to the multivariate case we are dealing with extremely small numbers ($\ll \exp(-30)$) and for this reason the method may be very susceptible to differences in computational precision in this case. The method held up well for the wind/whitecap multivariate extension (which involved the inclusion of additional X's) but has not performed well in this case with the inclusion of additional Y's. This may be because the inclusion of additional Y's increases the dimension of the regression density function, whereas the inclusion of additional X's does not.

In fact, in view of the results presented in later sections, a number of aspects of the analysis of this data set are not at all surprising. Firstly, since the data indicate a very strong underlying correlation, it is to be expected that the classical estimator will perform well. Secondly, for the same reason, we can expect the Lwin & Maritz type estimator to perform poorly.

7. THE SIMULATION STUDY

In Section 1, we evaluated the performance of a number of point and interval estimators of wind speed given whitecap coverage when applied to each of four data sets. The five estimators involved were:

- E1 (i) the inverse
- E2 (ii) the classical
- E3 (iii) estimated empirical Bayes
- E4 (iv) mean of predictive distribution
- E5 (v) mode of predicted distribution

together with corresponding interval estimators, each of which is defined in Section 3. The general conclusion drawn was that, with the exception of estimator (ii), which was considerably inferior, all the other estimators are broadly comparable in terms of predictive performance. This conclusion is supported by the results of several previous studies.

In this section we further evaluate the performance of these estimators by computer simulation. We concentrate in particular on the robustness of the estimators, and on the effect of sample size on the predictive ability of the estimators. The classical assumption is that both variables in the calibration study have normal distributions; this is the first situation we have studied. We have subsequently allowed for non-normal distributions for each variable in turn, and for both together. Another factor which has emerged as being of importance in determining the relative and absolute merits of the different estimators is the (true) correlation between the two variables, and the effect of this factor has also been examined.

This section is divided into two parts; the first, Section 8, is concerned with the point estimators, and the second, Section 9, with interval estimators.

8. COMPARISON OF POINT ESTIMATORS

8.1 The Point Estimators

The estimators being compared are the five referred to above with the following additions:

(a) Two alternative versions of estimator E3 [the Empirical Bayes estimator] are developed, viz.,

E6: assuming the errors follow a Student t-distribution, and estimating its variance in the standard manner and

E7: as in (i), except that we use a maximum likelihood estimate of the variance of the t-distribution.

(b) A further alternative version of estimator E3 is derived by assuming that the distributions of X and $Y|X$ are $N(\mu_x, \sigma_x^2)$ and $N(\alpha + \beta X, \sigma_y^2)$, respectively. Then, by straightforward probability calculus we have

$$f(x|y) \sim N\left\{\frac{\beta(y-\alpha)\sigma_x^2 + \mu_x\sigma_y^2}{\beta^2\sigma_x^2 + \sigma_y^2}, \frac{1}{\frac{\beta^2}{\sigma_y^2} + \frac{1}{\sigma_x^2}}\right\} \quad (8.1)$$

Hence an "empirical" Bayes estimator of X given Y is

$$\begin{aligned} \text{E8: } & \frac{\hat{\sigma}_x^2 \hat{\beta}(y - \hat{\alpha}) + \bar{x} \hat{\sigma}_y^2}{\hat{\beta}^2 \hat{\sigma}_x^2 + \hat{\sigma}_y^2} \end{aligned} \quad (8.2)$$

Note that as $\sigma_y^2/\sigma_x^2 \rightarrow 0$ (i.e., $\rho^2 \rightarrow 1$, where ρ is the (true) correlation coefficient of X and Y), this estimator $\rightarrow (y - \hat{\alpha})/\hat{\beta}$ ($\hat{\beta} \neq 0$), the classical

estimator. Therefore, certainly for large samples, we would expect the performance of the classical estimator, which in general is not good, to improve as $\rho^2 \rightarrow 1$. Note further that (see Appendix A) the estimator E8 is virtually identical with E1 for any reasonably large sample size N, thereby providing justification for the use of the estimator E1.

8.2 Simulation

The criterion of comparison of the different estimators is their mean-squared error of prediction. The basic assumption is that we have two random variables X, Y such that

$$E\{y|x\} = \alpha + \beta X$$

(8.3)

$$V(y|x) = \sigma_y^2$$

The study involves a number of different assumptions concerning the form of the distributions of X and Y|X and these are detailed below. The (true) values of α and β are taken to be 0 and 1, respectively. An initial random sample of size n is generated from which the predictive relation is derived. Then 100 further pairs of observations were simulated from the same true model, and a prediction of the x-value corresponding to each y-value is made using each of the eight estimators.

The above exercise was carried out 2000 times for every combination of the following parameters:

Sample size N	20,	40,	80
Squared corr. coefficient	.1,	.5,	.9

The overall exercise was repeated for each of the following combinations of assumptions regarding the forms of the distribution of X and Y:

- | | |
|-----------------------------|---|
| 1. X: $N(0,1)$ | Error: Normal, mean 0 |
| 2. X: $N(0,1)$ | Error: t 3 d.f. |
| 3. X: $N(0,1)$ | Error: Stretched Normal (Gaver
(1982)) |
| 4. X: t, 3 d.f., variance 1 | Error: Normal, mean 0 |
| 5. X: t, 3 d.f., variance 1 | Error: t, 3 d.f. |

The error variance in each case was fixed so as to give the required correlation between X and Y.

The results arising from each series of assumptions are presented in Tables 8.1 through 8.5, respectively.

TABLE 8.1

MSE of Various Predictors

		X: Normal	Error: Normal	
ρ -squared	Estimator	N = 20	N = 40	N = 80
.1	E1	1.02	0.95	0.92
	E2	149.50	2063.04	174.14
	E3	1.02	0.95	0.92
	E4	1.01	0.95	0.92
	E5	1.01	0.95	0.92
	E6	1.07	1.00	0.96
	E7	1.01	0.96	0.92
	E8	1.01	0.95	0.92
.5	E1	0.56	0.52	0.51
	E2	1.51	1.21	1.04
	E3	0.58	0.53	0.51
	E4	0.56	0.52	0.51
	E5	0.56	0.52	0.51
	E6	0.63	0.58	0.56
	E7	0.62	0.56	0.53
	E8	0.56	0.52	0.51
.9	E1	0.11	0.10	0.10
	E2	0.13	0.12	0.11
	E3	0.15	0.12	0.11
	E4	0.11	0.10	0.10
	E5	0.11	0.10	0.10
	E6	0.18	0.13	0.13
	E7	0.18	0.13	0.13
	E8	0.11	0.10	0.10

TABLE 8.2

MSE of Various Predictors

X: Normal Error: t, 3 d.f.

ρ -squared	Estimator	N = 20	N = 40	N = 80
.1	E1	1.04	0.95	0.98
	E2	95.13	141.63	19.67
	E3	0.97	0.92	0.90
	E4	1.00	0.95	0.97
	E5	0.99	0.95	0.95
	E6	0.96	0.88	0.85
	E7	0.93	0.87	0.83
	E8	1.03	0.95	0.98
.5	E1	0.57	0.49	0.51
	E2	1.32	1.02	1.03
	E3	0.51	0.48	0.48
	E4	0.56	0.49	0.51
	E5	0.55	0.49	0.50
	E6	0.50	0.47	0.44
	E7	0.50	0.46	0.43
	E8	0.56	0.49	0.51
.9	E1	0.11	0.11	0.12
	E2	0.12	0.12	0.13
	E3	0.14	0.11	0.11
	E4	0.11	0.11	0.12
	E5	0.12	0.11	0.12
	E6	0.16	0.12	0.11
	E7	0.17	0.12	0.11
	E8	0.11	0.11	0.12

TABLE 8.3

MSE of Various Predictors

X: Normal Error: Stretched Normal

ρ -squared	Estimator	N = 20	N = 40	N = 80
.1	E1	1.00	0.97	0.92
	E2	3230.46	66.92	14.31
	E3	0.97	0.95	0.91
	E4	0.98	0.97	0.92
	E5	0.99	0.92	0.92
	E6	0.96	0.93	0.88
	E7	0.95	0.92	0.87
	E8	1.00	0.97	0.92
.5	E1	0.56	0.56	0.52
	E2	1.30	1.19	1.07
	E3	0.52	0.52	0.48
	E4	0.55	0.55	0.52
	E5	0.55	0.56	0.52
	E6	0.52	0.50	0.46
	E7	0.51	0.49	0.46
	E8	0.56	0.56	0.52
.9	E1	0.11	0.11	0.10
	E2	0.13	0.12	0.11
	E3	0.15	0.11	0.10
	E4	0.11	0.11	0.10
	E5	0.11	0.11	0.10
	E6	0.17	0.12	0.13
	E7	0.17	0.13	0.11
	E8	0.11	0.11	0.10

TABLE 8.4

MSE of Various Predictors

X: t, 3 d.f. Error: Normal

p-squared	Estimator	N = 20	N = 40	N = 80
.1	E1	0.99	0.95	0.92
	E2	2919.72	0.99	0.97
	E3	0.92	0.93	0.97
	E4	0.98	0.95	0.92
	E5	0.98	0.95	0.93
	E6	0.96	1.00	0.98
	E7	0.92	0.95	0.96
	E8	0.98	0.95	0.92
.5	E1	0.58	0.55	0.54
	E2	1.82	1.14	1.18
	E3	0.61	0.58	0.56
	E4	0.59	0.55	0.54
	E5	0.59	0.57	0.55
	E6	0.72	0.70	0.68
	E7	0.69	0.69	0.66
	E8	0.59	0.55	0.54
.9	E1	0.11	0.10	0.10
	E2	0.13	0.11	0.11
	E3	0.21	0.18	0.16
	E4	0.11	0.11	0.10
	E5	0.11	0.12	0.10
	E6	0.35	0.33	0.32
	E7	0.35	0.34	0.33
	E8	0.12	0.11	0.10

TABLE 8.5

MSE of Various Predictors

X: t, 3 d.f. Y: t, 3 d.f.

p-squared	Estimator	N = 20	N = 40	N = 80
.1	E1	1.09	0.87	0.87
	E2	1652.77	1.00	0.84
	E3	1.06	0.89	0.79
	E4	1.06	0.86	0.78
	E5	1.05	0.86	0.77
	E6	1.05	0.85	0.76
	E7	1.01	0.83	0.73
	E8	1.09	0.88	0.78
.5	E1	0.56	0.55	0.52
	E2	2.51	1.26	0.97
	E3	0.56	0.55	0.53
	E4	0.57	0.55	0.52
	E5	0.58	0.54	0.53
	E6	0.66	0.61	0.63
	E7	0.65	0.60	0.62
	E8	0.57	0.55	0.52
.9	E1	0.13	0.11	0.12
	E2	0.13	0.12	0.12
	E3	0.23	0.16	0.17
	E4	0.13	0.11	0.12
	E5	0.13	0.11	0.12
	E6	0.59	0.50	0.47
	E7	0.58	0.50	0.46
	E8	0.13	0.10	0.12

8.3 Discussion of Simulation Results

The criterion used for comparison of estimators--the mean-squared error--is, of course, scale dependent, and therefore the only meaningful comparison between estimators is the percentage difference in mean squared error.

Looking first at Table 8.1 (X and the error both Normal), we see that E1, E4, E5 and E8 are virtually indistinguishable in terms of predictive performance. The Lwin & Maritz type procedures (E3, E6, and E7) are somewhat inferior particularly for small sample size and/or large ρ^2 . For example, for $N = 20$, $\rho^2 = .9$, the appropriate Lwin & Maritz estimator (E3) is approximately 36% worse than the four "good" estimators in terms of mean squared error. The classical estimator (E2) turns out to be just as bad as might be expected from previous studies, although it does, as we predicted it should, appear to improve as ρ^2 increases.

In Table 8.2 (X Normal, error having a t-distribution), the four estimators E1, E4, E5 and E8 are again essentially identical in their performance. The classical estimator is again poor, with the same proviso as above. However, the Lwin & Maritz type estimators (E5, E6, E7) now perform very well, except for a combination of small N and large ρ^2 . The superiority of the most appropriate (and best) of these estimators (E7) is of the order of 10 to 15 percent reduction in mean squared error for ρ^2 in the range .1 to .5, although for larger ρ^2 and small sample sizes this difference is smaller, and in one case is reversed ($N = 20$, $\rho^2 = .9$). This is a general pattern that has emerged: The Lwin & Maritz type estimators do

not perform well for large ρ^2 , particularly when the corresponding sample sizes are small.

In Table 8.3 (X still Normal, the error having even longer, more straggling tails), the pattern is very similar to that of Table 7.2. Once again the estimators E1, E4, E5, E8 are broadly comparable. E2 is poor, and estimators E3, E6 and E7 are, with the type of exception mentioned above, superior (involving a reduction of up to about 12% in mean squared error).

In the remaining tables, we allow X to be non-Normal. In the case of Table 8.4 (X, t distribution, error Normal), estimators E1, E4, E5, and E8 are still virtually identical. E2 is still the worst, but E3 (which one would expect, given its definition, to be good in this case) is superior only for small ρ^2 , and this superiority is most marked for small ρ^2 , and this superiority is most marked for small N. For moderate ρ^2 (.5), E3 is marginally worse than E1, E4, E5, and E8, and for large ρ^2 , E3 is distinctly inferior. E6 and E7 are, in general, as might be expected in this case, very poor in their predictive capacity.

Finally, in Table 8.5 we have the case where neither X nor the error is Normally distributed. The pattern of Table 7.6 continues here, except that the cases where E3 is superior are even more limited, and the inferiority of E6 and E7 for large ρ^2 even more pronounced.

Some general conclusions can now be drawn from the combined results:

1. For all ρ^2 , and all N, regardless of underlying distributions, the estimators E1, E5, E6, and E8 are indistinguishable in terms of predictive performance, when that performance is measured in terms of MSE.

2. When both X and the error are Normal, one of the estimators $E1$, $E4$, $E5$, and $E8$ should be used. The Lewin & Maritz type estimator can be inferior in this case, particularly for large ρ^2 and small N .

3. When X is Normal, but the error is not, the LM estimators can be superior, except when there is a combination of high ρ^2 , and small N . Modifying the LM estimator to take account of the form of the error distribution ($E6$, $E7$) does lead to further reduction of the mean-squared error.

4. When X is long-tailed non-Normal, the range of superiority of the LM estimators is very limited--in fact it only occurs for small ρ^2 , and is most marked for small N . Calibration is probably not a very appropriate technique in that situation. Therefore, when X is non-Normal, one should probably utilize one of the estimators $E1$, $E4$, $E5$ or $E8$.

5. The estimator $E8$, which has not been studied before, performs very well in general. The inverse estimator $E1$ performs equally well, but estimator $E8$ has some appealing properties, viz.

- (i) it can be derived directly from our assumptions (7.3) and
- (ii) it leads to a simple and reliable confidence interval (cf. Section 8)
- (iii) Simple algebra (Appendix A) will show that $E8$ is essentially almost identical with $E1$, thus providing justification for use of $E1$.

6. Estimators $E4$ and $E5$ also perform well, but are computationally more difficult to obtain, and do not yield easily computable confidence

intervals. However, they do give a readily computable predictive distribution.

7. Sample size is not a major factor in the absolute size of the mean-squared error. Reading across any of the rows of any of the tables 8.1 through 8.5, we see relatively little reduction in MSE as we go from $N = 20$ to $N = 40$ to $N = 80$. The reduction is certainly small compared to the reduction as we go from $\rho^2 = .1$ to $\rho^2 = .5$ to $\rho^2 = .9$. This is not, of course, very surprising: it merely indicates that the main determining factor in the predictive capacity of the various calibration estimators is the strength of the actual (linear) relationship between the relevant variables. Nevertheless, certain ways of processing the data can have considerable advantages.

9. COMPARISON OF INTERVAL ESTIMATORS

9.1 The Interval Estimators

Although numerous point estimators have been derived and studied in connection with the calibration problem, the study of the interval estimation problem has been much less extensive. In this section we examine, again by means of simulation, the performance of a number of interval estimators. These estimators are as follows.

1. For the point estimator $E1$, we use the standard 95% confidence interval for the predicted value of X , given $y = y_0$, viz.

$$E1 \pm \left(1 + \frac{1}{N} + \frac{(\bar{y} - y_0)^2}{S_{yy}}\right)^{1/2} \times s \times t_{0.025}$$

where

$$S^2 = \frac{S_{xx} - \hat{\beta}^2 S_{yy}}{N-2}$$

2. Brownlee (1965) has suggested a 95% confidence interval related to the approach of point estimator E2. This interval, referred to in the following tables as the classical interval, has the disadvantage that it fails to exist in certain circumstances. Its performance is examined.

3. An empirical Bayes-type confidence interval based on the derivation of E8 is given by

$$E8 \pm t_{0.025} \frac{1}{\frac{\hat{\beta}^2}{\sigma_y^2} + \frac{1}{\sigma_x^2}},$$

and described herein as empirical Bayes type 1.

4. Lwin & Maritz have an alternative suggested procedure for deriving an empirical Bayes confidence interval. Three different intervals of this type are calculated, viz.,

(i) an interval based on assuming a normal distribution for the error term, and denoted by type 2;

(ii) an interval based on assuming a t-distribution for the error, and estimating its variance by the standard method--a type 3 empirical Bayes interval;

(iii) similar to (ii), except that a maximum likelihood approach is used to estimate the variance. This we call a type 4 interval.

All these intervals have the property (see Lwin & Maritz (1980)) that they can be semi-infinite. As such intervals make the calculation of average interval length impossible, and as their occurrence is rare, we omit them from our calculations, and merely record their frequency of occurrence.

5. It is possible to construct confidence intervals based on the predictive distribution of Aitchison & Dunsmore, but since this involves, for a single y -value, repeated numerical integration it is omitted from the simulation study.

9.2 Design of the Simulation Study

The design of the simulation study is identical with that in Section 7, except that, due to the omission of the (computationally lengthy) Aitchison & Dunsmore estimators E4 and E5, we are able to greatly expand the number of replications at each setting of ρ^2 , N . In fact we now repeat the experiment (of generating a sample, and 100 additional pairs of observations for prediction based on the sample) 2000 instead of 100 times. This means that for each ρ^2 , N configuration, we are constructing 200,000 confidence intervals. The intervals so constructed are compared in terms of percent coverage and average length.

The results are presented in Tables 9.1 through 9.5; these tables have a direct correspondence with Tables 8.1 through 8.5.

TABLE 9.1

Performance of Various Confidence Intervals

X: Normal

Error: Normal

Confidence Interval

ρ^2	Sample Size	Inverse		Emp. Bayes		Classical		% exist
		% Cov.	Av. Length	% Cov.	Av. Length	% Cov.	Av. Length	
.1	20	94.8	4.0	93.1	3.7	94.6	35.6	66.6 (22.7)
	40	94.9	3.8	94.1	3.7	96.6	37.0	83.2 (12.1)
	80	94.9	3.7	94.5	3.7	97.3	32.4	96.0 (2.9)
.5	20	94.8	3.0	93.6	2.8	96.1	6.2	(99.6) (0.4)
	40	95.1	2.9	94.4	2.8	95.7	4.4	100.0 0
	80	94.7	2.8	94.3	2.7	95.0	4.1	100.0 0
.9	20	94.9	1.3	93.5	1.3	95.0	1.4	100.0 0
	40	95.0	1.3	94.4	1.2	95.1	1.3	100.0 0
	80	94.9	1.2	94.6	1.2	94.9	1.3	100.0

TABLE 9.1 (continued)

ρ squared		Emp. Bayes 2			Emp. Bayes 3			Emp. Bayes 4		
Sample Size		% Cov.	Av. Length	% Exist	% Cov.	Av. Length	% Exist	% Cov.	Av. Length	% Exist
.1	20	86.9	3.2	100.0	88.3	3.5	100.0	88.5	3.5	100.0
	40	90.6	3.5	100.0	92.2	3.7	100.0	92.4	3.8	100.0
	80	93.5	3.7	100.0	94.1	3.8	100.0	94.2	3.9	100.0
.5	20	82.7	2.4	100.0	78.2	2.2	100.0	87.2	2.5	100.0
	40	90.2	2.7	100.0	86.4	2.5	100.0	89.7	2.7	100.0
	80	92.6	2.7	100.0	89.0	2.5	100.0	92.3	2.8	100.0
.9	20	78.8	1.1	99.0	42.6	0.6	98.6	50.2	0.7	98.9
	40	86.3	1.1	99.8	45.5	0.5	99.4	55.5	0.6	99.6
	80	90.5	1.2	99.8	46.6	0.5	99.6	59.2 0	0.6	99.8

TABLE 9.2

Performance of Various Confidence Intervals

X: Normal

Error: t 3 d.f.

Confidence Interval

p-squared		Inverse		Emp. Bayes 1		Classical		% exist
Sample Size		% Cov.	Av. Length	% Cov.	Av. Length	% Cov.	Av. Length	
.1	20	94.5	4.0	92.4	3.7	93.9	39.1	69.7 (15.1)
	40	94.8	3.8	93.8	3.6	95.4	46.8	(5.9)
	80	94.9	3.7	94.3	3.6	95.9	26.1	95.5 (1.6)
.5	20	93.8	2.8	92.2	2.6	94.1	7.3	1.3 (1.5)
	40	95.1	2.8	94.1	2.7	94.7	4.6	99.7
	80	95.5	2.8	95.1	2.8	95.2	4.1	99.9
.9	20	93.4	1.2	92.3	1.1	93.3	1.3	99.9
	40	94.6	1.2	94.1	1.2	94.5	1.3	100.0
	80	95.2	1.2	95.0	1.2	95.1	1.3	100.0

Table 9.2 (continued)

ρ Squared		Emp. Bayes 2			Emp. Bayes 3			Emp. Bayes 4		
Sample Size		% Cov.	Av. Length	% Exist	% Cov.	Av. Length	% Exist	% Cov.	Av. Length	% Exist
.1	20	85.9	3.1	100.0	87.2	3.2	100.0	87.5	3.3	100.0
	40	91.9	3.5	100.0	93.5	3.8	100.0	93.7	3.9	100.0
	80	93.9	3.6	100.0	94.5	3.8	100.0	94.6	3.8	100.0
.5	20	82.2	2.2	99.8	75.0	1.9	100.0	77.9	2.1	100.0
	40	88.3	2.4	99.9	84.2	2.1	100.0	86.3	2.2	100.0
	80	91.7	2.5	99.9	45.3	0.6	98.9	48.1	0.6	98.9
.9	20	75.9	1.0	98.6	45.3	0.6	98.9	48.1	0.6	98.9
	40	85.9	1.1	99.5	50.5	0.5	99.3	54.3	0.5	99.4
	80	90.5	1.1	99.9	51.5	0.4	99.7	55.7	0.4	99.7

TABLE 9.3

Performance of Various Confidence Intervals

X: Normal

Error: Stretched Normal

Confidence Interval

ρ -squared		Inverse		Emp. Bayes 1		Classical		
Sample Size		% Cov.	Av. Length	% Cov.	Av. Length	% Cov.	Av. Length	% exist
.1	20	94.5	4.0	92.6	3.7	94.2	58.7	67.5 (17.6)
	40	94.7	3.8	93.8	3.7	95.3	36.7	18.6 (8.5)
	80	94.9	3.7	94.4	3.6	96.0	26.7	95.2 (2.0)
.5	20	93.8	2.9	92.2	2.7	94.1	6.3	99.1 (.5)
	40	94.4	2.7	93.4	2.7	94.1	4.3	99.8
	80	94.9	2.7	94.6	2.7	94.8	4.1	100.0
.9	20	93.0	1.2	91.9	1.1	93.0	1.3	100.0
	40	93.9	1.2	93.5	1.2	93.8	1.3	100.0
	80	94.6	1.2	94.3	1.2	94.5	1.3	100.0

Table 9.3 (continued)

ρ Squared		Emp. Bayes 2			Emp. Bayes 3			Emp. Bayes 4		
Sample Size		% Cov.	Av. Length	% Exist	% Cov.	Av. Length	% Exist	% Cov.	Av. Length	% Exist
.1	20	85.5	3.1	100.0	87.5	3.3	100.0	87.7	3.4	100.0
	40	90.7	3.4	100.0	92.6	3.7	100.0	93.8	3.7	100.0
	80	93.3	3.6	100.0	94.3	3.9	100.0	94.3	3.9	100.0
.5	20	82.3	2.3	99.9	76.9	2.1	99.9	79.4	2.2	100.0
	40	88.0	2.4	100.0	83.4	2.1	100.0	90.7	2.5	100.0
	80	91.9	2.6	100.0	89.1	2.3	100.0	90.7	2.5	100.0
.9	20	77.3	1.1	98.9	48.1	0.6	99.8	50.8	0.6	98.9
	40	86.5	1.2	99.5	51.4	0.5	99.2	55.9	0.6	99.4
	80	90.0	1.2	99.9	53.9	0.4	99.8	58.9	0.6	99.8

TABLE 9.4

Performance of Various Confidence Intervals

X: t

Error: Normal

Confidence Interval

ρ -squared		Inverse		Emp. Bayes		Classical		% exist
Sample Size		% Cov.	Av. Length	% Cov.	Av. Length	% Cov.	Av. Length	
.1	20	93.5	3.7	92.1	3.5	94.1	53.1	63.2 (15.9)
	40	94.1	3.6	93.5	3.5	96.2	40.7	79.5 (13.3)
	80	94.5	3.6	94.2	3.5	97.1	43.0	92.5 (5.2)
.5	20	94.1	2.8	92.4	2.6	96.2	12.1	97.7 (2.5)
	40	96.2	2.7	93.4	2.6	95.9	5.3	100.0 (0.0)
	80	94.8	2.7	94.4	2.6	95.3	4.2	100.0
.9	20	94.4	1.3	92.9	1.2	94.9	1.4	100.0 (0.0)
	40	94.9	1.2	94.0	1.2	95.2	1.4	100.0 (0.0)
	80	94.7	1.2	94.3	1.2	95.0	1.3	100.0 (0.0)

Table 9.4 (continued)

ρ Squared		Emp. Bayes 2			Emp. Bayes 3			Emp. Bayes 4		
Sample Size		% Cov.	Av. Length	% Exist	% Cov.	Av. Length	% Exist	% Cov.	Av. Length	% Exist
.1	20	88.5	3.3	100.0	90.3	3.7	100.0	90.7	3.8	100.0
	40	91.4	3.3	100.0	92.5	3.7	100.0	92.6	3.9	100.0
	80	93.6	3.5	100.0	94.3	3.6	100.0	94.2	3.6	100.0
.5	20	83.9	2.2	99.8	78.4	2.0	100.0	81.4	2.2	100.0
	40	90.4	2.4	99.9	86.9	2.2	100.0	89.5	2.4	100.0
	80	92.0	2.5	99.9	88.6	2.3	100.0	91.4	2.5	100.0
.9	20	78.8	1.0	98.4	43.8	0.6	99.0	51.4	0.6	99.2
	40	86.0	1.1	99.0	45.3	0.6	99.4	56.2	0.7	99.5
	80	90.7	1.2	99.6	47.1	0.5	99.8	60.6	0.7	99.2

TABLE 9.5

Performance of Various Confidence Intervals

X: t

Error: t

Confidence Interval

ρ -squared		Inverse		Emp. Bayes		Classical		
Sample Size		% Cov.	Av. Length	% Cov.	Av. Length	% Cov.	Av. Length	% exist
.1	20	93.2	3.6	91.5	3.3	94.3	40.1	66.2 (16.1)
	40	94.0	3.5	93.1	3.4	95.4	141.0	79.0 (9.1)
	80	94.4	3.5	94.0	3.5	95.8	27.2	93.5 (2.5)
.5	20	92.7	2.6	90.9	2.4	94.0	7.0	96.6 (1.2)
	40	94.0	2.6	92.7	2.5	94.5	4.5	99.2 (0.1)
	80	94.0	2.6	93.6	2.5	94.7	4.0	99.7 (0.0)
.9	20	93.1	1.2	92.7	1.1	93.5	1.3	99.9 (0.0)
	40	94.0	1.1	93.2	1.1	94.1	1.3	99.9 (0.0)
	80	94.4	1.1	94.1	1.1	94.6	1.3	100.0 (0.0)

Table 9.5 (continued)

ρ Squared		Emp. Bayes 2			Emp. Bayes 3			Emp. Bayes 4		
Sample Size		% Cov.	Av. Length	% Exist	% Cov.	Av. Length	% Exist	% Cov.	Av. Length	% Exist
.1	20	86.7	3.0	99.9	88.4	3.4	100.0	88.7	3.4	100.0
	40	92.3	3.5	99.9	93.6	3.8	100.0	93.7	3.9	100.0
	80	93.6	3.5	99.9	94.2	3.6	100.0	94.2	3.6	100.0
.5	20	83.1	2.1	99.5	75.8	1.8	100.0	79.1	1.9	100.0
	40	89.3	2.2	99.8	84.9	2.0	100.0	86.9	2.0	100.0
	80	91.6	2.3	99.8	88.6	2.0	100.0	90.2	2.1	100.0
.9	20	76.8	0.9	98.3	44.2	0.6	99.1	47.8	0.6	99.2
	40	86.9	1.1	99.0	51.7	0.6	99.5	55.3	0.6	99.6
	80	90.6	1.1	99.3	55.5	0.5	99.7	59.2	0.5	99.7

9.3 Discussion of the Results

We discuss each of the interval types separately.

9.3.1 The Inverse

This interval performs extremely well, both in terms of % coverage and average interval length. Of the intervals studied it has uniformly the shortest average length for a given level of coverage. Its robustness in terms of coverage is very good. The actual coverage does not fall below 93% in any of the five distributional situations considered, and for sample sizes 40 and 80 it does not fall below 94%. For situations where X is Normal, the coverage is very close to 95%.

9.3.2 The Empirical Bayes type 1

In terms of % coverage and average length, this empirical Bayes interval has a performance profile very similar to that of the inverse. Its coverage, in general, tends to be somewhat lower than the required 95%, and the average length tends to be marginally less than that for the inverse. Its robustness is very similar to that described in relation to the inverse.

9.3.3 The Classical

This interval performs, in general, very poorly. In the first instance, we examine the situations where it fails to exist. The final column in each table gives the percentage of simulations for which this interval existed. In general, no real interval existed for a large percentage of the simulations when ρ^2 was small (.1) and particularly so if N was also small. The % of non-existing intervals decreased rapidly (from c. 30-35% to 4-5%) as N increased from 20 to 80. A further difficulty also

emerged, even in cases where the interval existed: in some such cases, the lower interval and point given by Brownlee was larger than the upper end-point. The percentage of such points is given in parentheses underneath the % existence figures in each table. Once again, the problem arises predominantly in a small ρ^2 , small N situation. To overcome this difficulty, we interchanged the end-points when this situation arose. Having made this adjustment, the interval does indeed give a % coverage close to 95%. However, in terms of interval length, it performs very poorly relative to the other intervals, with one exception. As ρ^2 becomes large ($\rho^2 = .9$), the average length tends to that of the other intervals. An explanation of this behavior is provided by the fact that as ρ^2 becomes large ($\sigma_y^2/\sigma_x^2 \rightarrow 0$ in our model) the center point of the Brownlee interval, viz.

$$\bar{x} + \frac{\hat{\beta}(y - \hat{\alpha})}{\hat{\beta}^2 - t_{0.025}^2 \times S^2 / \sum (x_i - \bar{x})^2}$$

which in general (if we consider the average lengths of the 95% confidence interval) is not a very good estimator, tends to

$$\bar{x} + \frac{y - \hat{\alpha}}{\hat{\beta}}$$

i.e., the classical estimator, E2. We have already seen that there is reason to expect this estimator to be good for large ρ^2 .

9.3.4 The Empirical Bayes (types 2-4)

In general, these intervals do not perform well, particularly in terms of coverage. The coverage is close to 95% only for the case of small ρ^2 combined with large N. Otherwise the coverage is less than 95%, and in some cases (particularly for large ρ^2) very much less than 95%. As we have previously noted in the simulation study of point estimators, the corresponding point estimators also perform very poorly for the same range of parameter values. Intervals of this type are not to be recommended.

9.3.5 General Conclusions

Of the six different intervals studied, that associated with the inverse point estimator is uniformly the best. It is by far the most robust to departures from underlying assumptions, and is strongly recommended for use in construction of confidence intervals for the calibration problem.

10. ACKNOWLEDGMENT

The research of one author (IGO'M) was facilitated by his tenure, at the Naval Postgraduate School, Monterey, California, of a Research Associateship administered by the National Research Council. D.P. Gaver acknowledges the partial support of the Probability and Statistics Program of the Office of Naval Research, Washington D.C.

11. REFERENCES

- Aitchison, J. and I.R. Dunsmore (1975), Statistical Prediction Analysis. Cambridge University Press, London.
- Brown, P.J. (1982). Multivariate Calibration. J. Royal Statist. Soc., B, 44, No. 3, pp., 287-321.

- Brownlee, K.A. Theory and Methodology in Sciences and Engineering. Wiley, London.
- Gaver, D.P. (1982). Stochastic Modeling: Ideas and Techniques. In: Probability Theory and Computer Science by G. Latouche and G. Louchard (Eds.). Academic Press, London.
- Krutchkoff, R.G. (1967). Classical and Inverse Regression Methods of Calibration. Technometrics, 9, pp. 425-439.
- Krutchkoff, R.G. (1969). Classical and Inverse Regression Methods of Calibration in Extrapolation. Technometrics, 11, pp. 605-608.
- Lwin, T. and J.S. Maritz (1980). A Note on the Problem of Statistical Calibration. App., Statist., 29, No. 2, pp. 135-141.
- Mardia, K.V., J.T. Kent and J.M. Bibby (1979). Multivariate Analysis. Academic Press, London.
- Monahan, E.C. (1971). Oceanic Whitecaps. Jour. Phys. Oceanogr., 1, pp. 139-144.
- Monahan, E.C. and I.G. O'Muircheartaigh (1980). Optimal Power-Law Description of Oceanic Whitecap Coverage Dependence on Wind Speed. Jour. Phys. Oceanogr., 10, pp. 2094-2099.
- Monahan, E.C., P.A. Bawyer, D.M. Doyle, M.P. Fitzgerald, I.G. O'Muircheartaigh, M.C. Spillane and J.J. Taper. (1981). Whitecaps and the Marine Atmosphere. Tech. Rep., 3, 125 pp. Univ. College, Galway, Ireland.
- Toba, Y. and M. Chaen. (1973). Quantitative Expression of the Breaking of Wind Waves on the Sea Surface. Records Oceanogr. Works Japan, 12, pp. 1-11.
- Wu, J. (1979). Oceanic Whitecaps and Sea State. Jour. Phys. Oceanogr., 9, pp. 1064-1068.
- Williams, E.J. (1969). A Note on Regression Methods in Calibration. Technometrics, 11, pp. 189-192.

Appendix

Given $y = y_0$, if we define E1 to be

$$E1 = \hat{\alpha}^* + \hat{\beta}^* y_0 \quad (A.1)$$

where

$$\hat{\beta}^* = \frac{S_{xy}}{S_{yy}}, \quad \text{and} \quad \hat{\alpha}^* = \bar{x} - \hat{\beta}^* \bar{y},$$

using conventional notation, and E8 to be

$$E8 = \frac{\hat{\sigma}_x^2 \hat{\beta}^2 (y_0 - \hat{\alpha}) + \bar{x} \hat{\sigma}_y^2}{\hat{\beta}^2 \hat{\sigma}_x^2 + \hat{\sigma}_y^2} \quad (A.2)$$

where $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$, $\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$, and $\hat{\sigma}_x^2 = \frac{S_{xx}}{n-1}$, $\hat{\sigma}_y^2 = \frac{S_{yy} - \hat{\beta}^2 S_{xx}}{n-2}$

then

$$E8 = \frac{\frac{S_{xx}}{n-1} \frac{S_{xy}}{S_{xx}} (y_0 - [\bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}]) + \bar{x} \frac{S_{yy} - \hat{\beta}^2 S_{xx}}{n-2}}{\hat{\beta}^2 \frac{S_{xx}}{n-1} + \frac{S_{yy} - \hat{\beta}^2 S_{xx}}{n-2}}$$

and if n is reasonably large, so that $n-1 \approx n-2$, then

$$E8 \approx \frac{s_{xy} \left[y_0 - \left\{ \bar{y} - \frac{s_{xy}}{s_{xx}} \bar{x} \right\} \right] + \left\{ s_{yy} - \hat{\beta}^2 s_{xx} \right\}}{\hat{\beta}^2 s_{xx} + s_{yy} - \hat{\beta}^2 s_{xy}}$$

$$= \frac{s_{xy}}{s_{yy}} y_0 - \frac{s_{xy}}{s_{yy}} \bar{y} + \frac{s_{xy}^2 \bar{x}}{s_{xx} s_{yy}} + \bar{x} - \frac{s_{xy}^2}{s_{xx} s_{yy}} \bar{x}$$

$$= \left(\bar{x} - \frac{s_{xy}}{s_{yy}} \bar{y} \right) + \frac{s_{xy}}{s_{yy}} y_0$$

$$= \hat{\alpha}^* + \hat{\beta}^* y_0 = E1 .$$

NO. OF COPIES

Dr. D. F. Daley
 Statistics Dept. (I.A.S)
 Australian National University
 Canberra A.C.T. 2606
 AUSTRALIA

1

Prof. F. J. Anscombe
 Department of Statistics
 Yale University, Box 2179
 New Haven, CT 06520

1

Dr. David Brillinger
 Statistics Department
 University of California
 Berkeley, CA 94720

1

Dr. R. Gnanadesikan
 Bell Core
 435 South Street
 Morris Township, NJ 07960

1

Prof. Bernard Harris
 Dept. of Statistics
 University of Wisconsin
 610 Walnut Street
 Madison, WI 53706

1

Dr. D. R. Cox
 Department of Mathematics
 Imperial College
 London SW7
 ENGLAND

1

Dr. A. J. Lawrance
 Dept. of Mathematical Statistics
 University of Birmingham
 P. O. Box 363
 Birmingham B15 2TT
 ENGLAND

1

Professor W. M. Hinich
 University of Texas
 Austin, TX 78712

1

Dr. John Copas
 Dept. of Mathematical Statistics
 University of Birmingham
 P. O. Box 363
 Birmingham B15 2TT
 ENGLAND

1

P. Heidelberger IBM Research Laboratory Yorktown Heights New York, NY 10598	1
Prof. M. Leadbetter Department of Statistics University of North Carolina Chapel Hill, NC 27514	1
Dr. D. L. Iglehart Dept. of Operations Research Stanford University Stanford, CA 94350	1
Dr. D. Vere-Jones Dept. of Mathematics Victoria University of Wellington P. O. Box 196 Wellington NEW ZEALAND	1
Prof. J. B. Kadane Dept. of Statistics Carnegie-Mellon Pittsburgh, PA 15213	1
Prof. J. Lehoczky Department of Statistics Carnegie-Mellon University Pittsburgh, PA 15213	1
Dr. J. Maar (R51) National Security Agency Fort Meade, MD 20755	1
Dr. M. Mazumdar Dept. of Industrial Engineering University of Pittsburgh Pittsburgh, PA 15235	1
Prof. M. Rosenblatt Department of Mathematics University of California-San Diego La Jolla, CA 92093	1
Prof. Rupert G. Miller, Jr. Statistics Department Sequoia Hall Stanford University Stanford, CA 94305	1

Prof. I. R. Savage Dept. of Statistics Yale University New Haven, CT 06520	1
Dr. Paul Shaman National Science Foundation Mathematical Sciences Section 1800 G. Street, NW Washington, D. C. 20550	1
Prof. W. R. Schucany Dept. of Statistics Southern Methodist University Dallas, TX 75222	1
Prof. D. C. Siegmund Dept. of Statistics Sequoia Hall Stanford University Stanford, CA 94305	1
Prof. H. Solomon Department of Statistics Sequoia Hall Stanford University Stanford, CA 94305	1
Dr. Ed Wegman Statistics & Probability Program Code 411(SP) Office of Naval Research Arlington, VA 22217	1
Dr. P. Welch IBM Research Laboratory Yorktown Heights, NY 10598	1
Dr. Marvin Moss Office of Naval Research Arlington, VA 22217	1
Dr. Roy Welsch Sloan School M. I. T. Cambridge, MA 02139	1
Pat Welsh Head, Polar Oceanography Branch Code 332 Naval Ocean Research & Dev. Activity NSTL Station, MS 39529	1

Dr. Douglas de Priest Statistics & Probability Program Code 411(SP) Office of Naval Research Arlington, VA 22217	1
Dr. Morris DeGroot Statistics Department Carnegie-Mellon University Pittsburgh, PA 15235	1
Prof. J. R. Thompson Dept. of Mathematical Science Rice University Houston, TX 77001	1
Prof. J. W. Tukey Statistics Department Princeton University Princeton, NJ 08540	1
Dr. Daniel H. Wagner Station Square One Paoli, PA 19301	1
Dr. Colin Mallows Bell Telephone Laboratories Murray Hill, NJ 07974	1
Dr. D. Pregibon Bell Telephone Laboratories - AT&T Murray Hill, NJ 07974	1
Dr. Jon Kettenring Bell Core 435 South Street Morris Township, NJ 07960	1
Dr. David L. Wallace Statistics Dept. University of Chicago 5734 S. University Ave. Chicago, IL 60637	1
Dr. F. Mosteller Dept. of Statistics Harvard University Cambridge, MA 02138	1
Dr. S. R. Dalal Bell Laboratories - AT&T Mountain Avenue Murray Hill, NJ 07974	1

Prof. Donald P. Gaver Code 55Gv Naval Postgraduate School Monterey, California 93943-5100	20
Prof. Patricia Jacobs Code 55Jc Naval Postgraduate School Monterey, California 93943-5100	1
Dr. Guy Fayolle I.N.R.I.A. Dom de Voluceau-Rocquencourt 78150 Le Chesnay Cedex FRANCE	1
Dr. M. J. Fischer Defense Communications Agency 1860 Wiehle Avenue Reston, VA 22070	1
Prof. George S. Fishman Curr. in OR & Systems Analysis University of North Carolina Chapel Hill, NC 20742	1
Prof. Guy Latouche University Libre Bruxelles C. P. 212 Blvd De Triomphe B-1050 Bruxelles BELGIUM	1
Library Code 1424 Naval Postgraduate School Monterey, CA 93943-5100	4
Dr. Alan F. Petty Code 7930 Naval Research Laboratory Washington, DC 20375	1
Prof. Bradley Efron Statistics Department Sequoia Hall Stanford University Stanford, CA 94305	1
Prof. Carl N. Morris Dept. of Mathematics University of Texas Austin, TX 78712	1

Dr. John E. Rolph RAND Corporation 1700 Main Street Santa Monica, CA 90406	1
Prof. Linda V. Green Graduate School of Business Columbia University New York, NY 10027	1
Dr. David Burman Bell Laboratories - AT&T Mountain Avenue Murray Hill, NJ 07974	1
Dr. Ed Coffman Bell Laboratories - AT&T Mountain Avenue Murray Hill, NJ 07974	1
Prof. William Jewell Operations Research Department University of California, Berkeley Berkeley, CA 94720	1
Dr. Tom A. Louis Biostatistics Department Harvard School of Public Health 677 Huntington Avenue Boston, MA 02115	1
Dr. Nan Laird Biostatistics Department Harvard School of Public Health 677 Huntington Avenue Boston, MA 02115	1
Dr. Marvin Zelen Biostatistics Department Harvard School of Public Health 677 Huntington Avenue Boston, MA 02115	1
Dr. John Orav Biostatistics Department Harvard School of Public Health 677 Huntington Avenue Boston, MA 02115	1

Prof. R. Douglas Martin Chairman Department of Statistics, GN-22 University of Washington Seattle, WA 98195	1
Prof. W. Stuetzle Department of Statistics University of Washington Seattle, WA 98195	1
Prof. F. W. Mosteller Department of Statistics Harvard University 1 Oxford Street Cambridge, MA 02138	1
Dr. D. C. Hoaglin Department of Statistics Harvard University 1 Oxford Street Cambridge, MA 02138	1
Prof. N. D. Singpurwalla George Washington University Washington, D. C. 20052	1
Center for Naval Analyses 2000 Beauregard Street Alexandria, VA 22311	1
Prof. H. Chernoff Department of Mathematics M. I. T. Cambridge, MA 02139	1
Dr. T. J. Ott Bell Core 435 South Street Morris Township, NJ 07960	1
Alan Weiss AT&T Bell Laboratories Murray Hill, NJ	1
Operations Research Center, Room E40-164 Massachusetts Institute of Technology Attn: R. C. Larson and J. F. Shapiro Cambridge, MA 02139	1

Research Administration Office
Code 012
Naval Postgraduate School
Monterey, CA 93943

1

DUDLEY KNOX LIBRARY



3 2768 00347546 8